DOCUMENT RESUME

ED 319 358 IR 014 209

AUTHOR Tucker, Richard N.
TITLE Comparative Study on Criteria and Procedures for the Evaluation of Educational Software. Summary.
SPONS AGENCY United Nations Educational, Scientific, and Cultural Organization, Paris (France).
PUB DATE Apr 89
NOTE 8p.; Paper presented at a Conference on Education and Informatics: Strengthening International Cooperation (Paris, France, April 12-21, 1989).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; *Computer Assisted Instruction; *Computer Software; Courseware; *Evaluation Criteria; *Evaluation Methods; Foreign Countries; *Microcomputers; Questionnaires; Surveys
IDENTIFIERS International Council for Educational Media

ABSTRACT

The International Council for Educational Media (ICEM) was asked to conduct a study based on reports from a selected number of the ICEM member countries representing different styles of educational organization (from centralized to decentralized), different organizational methods for the production of educational software (from governmental to commercial), and different levels of integration of the computer into education. A detailed questionnaire, which was developed as the survey instrument, requested information about professional, applications, and didactic software packages from software developers, converters, and users in Canada, England, Hungary, Italy, Japan, the Netherlands, Scotland, and the United States. Specific questionnaire items addressed issues such as the educational context in which the software was used, the nature of software development (governmental or commercial), how decisions about content are made, criteria for field testing, production criteria, criteria for the conversion of software brought in from other countries, criteria for the selection of commercially available software, and criteria used to evaluate the use of software in the classroom. A summary of the results for each of these areas is provided and conclusions are discussed relating to software development, production, selection, conversion, and research. (GL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

EDUCATION AND INFORMATICS:
strengthening international cooperation
Paris, 12 to 21 April 1989

Background paper to
Theme 2: Strategies
Session 2.2

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Richard N Tucker
Director Educational and Information Technology
Netherlands Institute for Audio-visual Media

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Suzanne Adlung

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED319358

A Summary of a "Comparative Study on Criteria and Procedures for the Evaluation of Educational Software" undertaken for UNESCO.

Terms of Reference

Under the UNESCO program on enhancing the quality of educational software ICEM, the International Council for Educational Media, was asked to conduct a study based on reports from a selected number of the ICEM member countries. This selection was to represent different styles of educational organisation, from the centralised to the decentralised, and the different organisational methods for the production of educational software from governmental to commercial. The countries were also to be selected to represent different levels of integration of the computer into education. On this last point it was reasonable to select countries with a relatively high integration in order to find enough evidence of the evaluation of software

Types of program

a) Professional Programs. These are programs produced initially for the professional or business community and which may later be used for education, e.g. wordprocessors, databases, spreadsheets, graphics and CAD/CAM etc.

b) Open or Applications Programs.These are sometimes called "content-free" programs, although some are in fact built around a specific content which can be added to or altered by the user. Some are simplified versions of the commercial professional programs. The adaptation to the educational market often rests on the provision of different documentation and some in-built examples.

c) Didactic Programs. These are programs with specific content and usually with stated learning objectives or goals. Here can be found the bulk of the "educational" software including; drill-and-practice, tutorial, simulation etc.

Evaluation

Evaluation may take place during at least three phases. Although there can never be fixed boundaries between the categories these may be regarded as;

IR014209

BEST COPY AVAILABLE

a) Development

The **Producers** of programs must have criteria on which they base the decision to produce a program, technical and methodological criteria for the development, and user criteria against which they test the product.

b)Selection for Translation and Conversion

The needs of the **Converters** are seen mainly in the smaller countries which, because of market sizes, are more concerned with the purchase and conversion of existing software than developing their own.

c) Selection by Authorities and Schools

the eventual **Users** of programs must have criteria by which they determine which programs are bought to meet their educational needs. This is perhaps the most widely known form of software evaluation, particularly because of the publicity which it has received from the USA and Canada.

The questionnaire, which was designed as the survey instrument, attempted to provide a matrix within which the commissioned authors would write the national studies. It was recognised that not all the elements in the matrix would be applicable in every country.

Software / Evaluator	Professional	Applications	Didactic
Developers			
Converters			
Users			

The Questions

A detailed questionnaire was developed and given to each of the correspondents. This asked for background on the educational context since this is seen as major conditioning factor. The nature of software development, whether Government subsidised or commercial was covered together with supply and how the decisions about content are made. The criteria for any field testing were requested. Production criteria followed in order to examine the economic factors which influence decision making. Where software is bought in from other countries and converted in any way the study sought the criteria by which this was done.

From experience it was obvious that the majority of evaluation is related to the selection of commercially produced software. This needed to be examined to determine if there were common criteria. It was also important to discover whether the evaluation for selection which we have come to know from the USA and Canada operated in other countries.

Finally the questionnaire asked for criteria used to evaluate the use of programs in the classroom.

The Countries

Contributions to this study came from Canada (Alberta, Ontario and Quebec), England, Hungary, Italy, Japan, Netherlands, Scotland, USA.

Summary

1. The Contexts

Though evaluations, one might even say the evaluation industry, to be found in the USA are frequently cited in reports on the use of microcomputers, a number of specific aspects are clear: many people (outside the USA) seem to feel that they should also have institutions such as EPIE; it also seems that *very few people* outside the target area of the USA do anything about these American reviews, in that they *buy and convert very little foreign material*; and the contexts in which these distant viewers stand is often radically different from that of the USA.

America with its system of devolved school control, operates with a large body of commercial production originating outside the education structure. Thus the market place imposes a demand on education to produce a "best buy" list. Canada, also a federation, presents a wide range of provincial organisations. Though there is a national software review system based in Alberta not all provinces base their software selection so strongly on reviews of commercial products. Ontario chooses projects for development based on submitted proposals and then ensures widespread distribution of the finished products.

When one considers Europe, other patterns emerge. The UK development projects have ensured a high degree of hardware penetration into the schools. The software development is not so consistent. In Scotland such production is largely centralised whereas in the rest of the UK the products of the projects are sold off to commercial publishers. Another factor which must be taken into account when considering the UK schools is that not only is there a largely unified school system but also that these schools have been working in a context of resource-based, project-based, and even inter-disciplinary education for the past fifteen years at least. This has undoubted advantages for the integration of computers into education. The Dutch, Italian and Hungarian examples run the gamut of a highly complex system of schools in the Netherlands with an inbuilt freedom of choice, through the Italian situation where the school system is not complex but there is apparently a lack of central guidance for the provision of hardware or software, to the Hungarian state system where the decision making is perhaps the most centralised. Japan despite it thrusting image in the world of computers is currently taking the most cautious stance of all the countries reported on here. The penetration of information technology into the Japanese schools is less well advanced. The traditional nature of many of the attitudes in Japanese education provides a resistance to innovation which is also to be found in many other countries. It is however encouraging to read the report in the latest edition of Audio-visual Education in Japan, produced by the Ministry of Education, which presents a more positive picture than that which was available to the correspondent through the 1984 and 1986 figures.

What can be drawn from the descriptions is that it is not easy to separate software and the criteria for its evaluation from the context and history from which it has grown.

2. Development of Software

Wherever software is being developed it is with an eye on the market. Apart from a small amount of development for personal satisfaction or to solve a very local problem, software developers want others to use their programs. That market may range from a small group of teachers to whom that program is freely given and who pay with their gratitude and accolades, to a potential pool of many millions of people - e.g. the major language groupings of the world. What can be deduced from these reports and others which were previously available is that criteria for the evaluation of programs during design and development are most likely to be found where educational institutions play a central role. The Scottish SMDP and the French software development linked to "L'Informatique pous Tous" both show clearly the way in which proposals from teachers can be formulated within guidelines and against sets of standards. This ensures at least some of the aspects of the quality of the program before it is finally produced. Toronto, for those programs which fall within the provincial program, has equally stringent criteria.

Because the criteria and procedures for these projects have been widely published we have become

sensitised to the fact that we frequently know very little of how other programs are produced or evaluated. One has reason to believe that much of it, even the very best programs, is written by the sort of educationalist who previously wrote those excellent text books - out of personal and professional experience but without much field-testing and rarely against lists of educational criteria.

3 Production

The same sort of range of experience can be seen in the production stages of a program, from countries which are mainly commercially based to those in which central government or large institutions have a major role to play. In both cases economic factors have a higher priority than educational criteria. One has to assume that these educational criteria had full operation in the early stages of planning and development. Where national projects such as those to be found in the Netherlands or Hungary arise, decisions can be taken to make large scale provision of particular programs, which negates the normal market criteria. One has also to assume that these programs have been selected for sound educational reasons.

When producing for an open market the normal criteria for deciding on any product must operate. Is the market big enough? Will somebody (and not necessarily the customer) pay for the product? Large language markets such as the English speaking world provide a good test of the quality of a product. For many other countries such criteria cannot be applied. If the maximum potential market in a particular language area is insufficient to cover the development and production costs then either those costs have to be lowered, the selling price increased, or a cost- covering subsidy has to be found from somewhere.

One of the results of these pressures - in an attempt to cut costs - has been the selection of cheaper systems which later fail to match up to the needs of the users and the marketting of programs which may create negative reactions to the computer over a period. One can cite here the underuse bordering on rejection of the computer which has been reported from the USA and attributed by some writers to the plethora of dull boring (but cheap!) drill-and-practice programs. In this sort of situation the publishers, with their large overhead costs may be reluctant to enter the market and the future may lie with the small unit which can manage to produce high quality software.

4 Conversion

Of the all the contributing countries only the Netherlands has made any consistant effort in converting software originally produced in other countries. It is therefore not justifiable to attempt to draw conclusions about criteria in this field. Whilst not underestimating the problems of converting programs from one computer system to another, from one language to another and from one cultural context to another, the Dutch have shown that there can be considerable benefits and cost savings if programs are selected carefully. The process is similar to the normal selection of software with the added dimensions of being able to determine if the way in which a program has been constructed makes is relatively easy to convert to another system. NIAM has discovered the value of having a few people who can build up their experience with specific systems. A further dimension is the ability to recognise which parts of a given program will need to be rewritten or replaced entirely with new examples in order to make the program acceptable culturally to the intended users. Underlying these questions are the essential elements to be found in the criteria for selection.

One must ask how much good material is not finding its way into other countries because it is written for one system which is not predominent in the other country. Equally one has to raise questions about instances where software is bought, say from the USA into the UK. How much consideration is given in that selection to items which are specifically American (context, terms of reference, even differences between the two types of English) which really ought to be converted if the program is to be used effectively. One suspects on the basis of these reports that the problem is not recognised. Conversion is something that is necessary for "foreign" programs. It would appear that smaller countries could do well to consider the criteria and processes of conversion.

5. Selection

Criteria for the selection of existing materials are perhaps the best known and most developed. They are most readily to be found where there are large amounts of software and the problems for the purchasers are those of knowledge of what is on the market and discrimination between the programs offered. It is not surprising therefore to find that in the USA there are a number of evaluation services. At a national level these may be published in journals, available to subscribers by post or even on-line. Some states have their own evaluation services and there are many more local or district services.

The selection of software is made in some cases by state bodies who then bulk-purchase on behalf of the schools, and in the majority of situation the selection is made by the individual teacher. In all cases it is important to the software producers that their products get evaluated. Without that sort of publicity (even with the risk of a bad review) making such a large audience aware of your products can be very expensive. The provision of information about what is on the market is obviously of major importance in a market such as the USA. Without this knowledge teachers will be hampered in their choice. Therefore many people demand, subscribe to and actively support evaluation services such as EPIE, MicroSIFT.

However some critical points emerge from the report. It seems that despite all these reviews the selection of media is more likely to result from having seen a demonstration than from reading a review and that even where a review is read, the general description carries more weight than the apparently more scientific evaluation table. Despite their inaccuracies and variable terminology publishers catalogues still function as one of the main sources of guidance for the selection of software.

There are also some problems with the instruments which have been devised for the evaluation services. They purport to be scientific, and present their evidence as though it were objective. Yet most of these evaluation forms were evolved for the recording of both fact and personal judgement about drill and practice programs. It is not now certain that they function as well when applied to the more sophisticated applications programs. If for the sake of argument considerable importance is given to the criterion which checks whether a program gives a cumulative score for the user, then this will have little relevance to a program which is built like a data-base. A largely graphic program may not match the criteria which were established when most programs were textual.

It is not simply a question of having a more flexible set of criteria. One has also to recognise that these reviews are applying, of necessity generalised criteria to specific programs which may be used in educational contexts in the same country which are very different from each other. Where these reports do seem to score highly is in their informative function and in the general description of the programs. Research has shown that it is the descriptive element of these reviews and journal articles which have a major influence on the way in which teacher select software.

The canadian reviews are also based on what look like objective reports. Yet when one considers the forms it is easy to see that apart from details like to size of the disc, the type of computer system, the language and the memory requirement, most of the answers required are a matter of judgement on the part of the reviewer. There are general checks as to whether it matches the curriculum, but against which curriculum is the program to be matched when there are so many to be found in a country as diverse as the USA or Canada. Local or state reviews may have less of a problem in this regard, where curricular decisions may make it easier to match programs to lesson requirements.

What one ends up with are generalised criteria which require subjective judgements (often on a 1 - 5 scale) from the reviewer. These are then presented in tabular form as though they are definite values. Such presentations have to be suspect. However both the US and Canadian experiences seem to point to a heavy reliance on the part of readers on the descriptive reviews. For these to have currency there has to be an agreed basis for comparison.
The Dutch review instrument relies heavily on subjective responses from the reviewers but recognises that these should be based on professional experience of the educational level and subject in question. It also provides quite clear guidance in its questioning. The resulting reports in SCEN are written like literary reviews. Readers can begin to get a feel for the way in which a particular journal reviews, just as they have their personal preferences for reviewers of football or theatre.

The strength and value of the descriptive review, particularly reviews of the application of a program - how it functions in a classroom - is strongly reinforced by the English report. Here there pointers towards a more fruitful use of software reviews. Those countries with less of a commercial market offer little evidence of there being any criteria for the selection of software. It could even be that once political decisions have been taken to produce particular items of software, that evaluation may expose the project to the risk of bad reviews. Policy may on occasion take precedence over pedagogical judgement.

6 Research

Given that so much attention is being paid to evaluation around the world. It seemed only right to ask whether it was of any value. Does it improve the choice of programs? Is a program which receives good reviews going to make a greater impact in the classroom? Given the vaste range of questions that are asked, if one bothered to put all the evaluation forms together, one would have expected to have found some research to back up all this effort. Sadly, but perhaps not surprisingly there is very little evidence of research. Some academic researches are reported from the USA. What is of interest an value is the comment which is made, explicitly or implicitly, by the American, Canadian, Dutch and English correspondents that out of all this reporting and evaluation what has proved to be the most valuable has been the illuminative research and evaluation. This is by its very nature more an evaluation of the software in use in a learning situation than an evaluation of the software as a "thing". The most prized evaluations are of the *process* rather than of the *product*.

CONCLUSIONS

1 Development: There is considerable value in having guidelines and principles for the development of software. These should be evolved with a strict relation to the existing curriculum and methods of learning. At the same time they should be designed in such a way as to guide program developers towards those methods and subjects which are desired for the near future. Where programs are developed within the educational structures, such guidelines should be applied as widely as the organisation of education permits. The only proviso is that those administering these standards should recognise that they are guidelines and not apply them so rigidly as to crush innovation. Where development takes place outside the educational system one can do little more than offer the guidelines. In smaller countries and language areas government support may have to be greater.

2. Production: After years of program production of very variable quality it ought to be possible within education to limit some of the idiosycracies. It ought to be possible to agree to a limited number of languages (if not a single language). It ought to be possible to agree to a small number of processes such as always keeping text in separate files so that translations or versions of a program can be made more easily. It ought to be possible to operate with a limited number of common commands. It ought to be possible to agree within any one system to a range of basic standards for screen presentation, e.g. number of lines of text on a screen, the design and spacing of questions, the use of colour and colour contrasts, and so forth. This should be possible to make it easier for the users to work with new software. There could even be agreement about the basic principles of putting menu bars and pull-down menus into programs. It ought to be possible.

3. Selection: The conclusion has to be drawn from these reports that where the schools have a choice of software there is an important role for what are now termed evaluations. It also has to be recognised that the majority of what are put forward as "evaluation forms" are very flawed and when applied rigorously across the range of software, largely inappropriate. Where evaluation forms do score highly is when they give the reviewer guidance on the aspects to look for and how to make judgements in the broad sense. They also act as a check list to reduce the risk that a reviewer might forget aspects which may be crucial to other potential buyers. If we recognise that the apparent objectivity of most of the published evaluations is an unsupportable pretence and that they are subjective but written on the basis of agreed standards, then we can give that

subjectivity its true value and start treating them as reviews of products. What we then need to do is extend this practice of reviewing to equally "guided" reviews of the programs in used - the process for which the software was originally written. Let us recognise that this illuminative evaluation of the process is the most valuable guidance that we can provide for education.

4. Conversion: Without labouring the argument that there are good reasons for converting software, criteria can be established by which the value of a program for possible conversion can be established. First of all use the criteria for selection with a special emphasis on the "curricular-fit".
If a program fits the curriculum, has a good track-record in its country of origin and has no effective competitors in the market into which it is being brought, then it is worth looking further at the programming itself. If this conforms to the ideal production criteria then there will be few difficulties. Judgements will have to be made on the degree of difficulty in taking a program from one system to another. The greatest thought and effort will have to be given to the cultural conversion - for which there can be few standard criteria.

5. Research: It appears doubtful if there is much valuable research to be done into the effectiveness of the present method of reviewing. With better review procedures there would be much to be gained from action research which supported the illuminative character of the reviews. The use of the software in classes is a far better and more valuable than comparative researches about the products, and certainly better than researches about how people fill in forms. Those searching for absolute answers with regard to any learning materials should be cautioned that no matter how carefully prepared a resource might be teachers and learners will use it in wholely different ways. Some of these ways will be better than that originally intended.

General remarks.

In concluding this report one is forced back to the consideration of why we want educational software to conform to a level of objective evaluation which we don't apply to any other medium. It is tempting to assume that the nature of the programs can be captured in tables of figures or circles in a matrix. Educational software at its best is powerful enough to resist being so caged. It comes to life in its use and the learning process often occurs between individuals in a small group of users of a program, stimulated by the program, rather than solely by one-to-one interaction with the program. We need to have more and better descriptions of the process of use in order to spread as many of the alternative practices and advantages throughout education.